Time-local discretization of fractional and related diffusive operators using Gaussian quadrature with applications

Florian Monteghetti [a,*], Denis Matignon [b], Estelle Piot [a]

[a] ONERA/DMPE, Université de Toulouse, 31055 Toulouse, France
[b] ISAE-SUPAERO, Université de Toulouse, 31055 Toulouse, France

Keywords:
Fractional derivative
Fractional calculus
Diffusive representation
Eigenvalue problems
Non-classical method
Completely monotone kernel

A B S T R A C T

This paper investigates the time-local discretization, using Gaussian quadrature, of a class of diffusive operators that includes fractional operators, for application in fractional differential equations and related eigenvalue problems. A discretization based on the Gauss–Legendre quadrature rule is analyzed both theoretically and numerically. Numerical comparisons with both optimization-based and quadrature-based methods highlight its applicability. In addition, it is shown, on the example of a fractional delay differential equation, that quadrature-based discretization methods are spectrally correct, i.e. that they yield an unpolluted and convergent approximation of the essential spectrum linked to the fractional derivative, by contrast with optimization-based methods that can yield polluted spectra whose convergence is difficult to assess.

1. Introduction

The broad focus of this article is the discretization of fractional operators using their so-called diffusive representation, for application in time-domain computations or eigenvalue problems.

The diffusive representation of fractional operators enables to recast them into an observer of an infinite-dimensional ODE: the long memory of the operator is reflected in the infinite dimension of the corresponding state space. Convolution operators that admit such a representation, known as diffusive operators, have a locally integrable completely monotone kernel. See [41,40] for definitions of fractional operators, [39,6] for an introduction to the class of diffusive operators, [21] for examples of diffusive operators, and [11,44] for a semigroup formulation of the state-space representation in the context of Volterra equations.

Provided that the diffusive representation is suitably discretized, it constitutes a time-local alternative to, for instance, fractional linear multistep methods [28] or methods based on the Grünwald–Letnikov approximation [42]. Existing discretization methods for the diffusive representation can be split into two categories: methods that rely on an optimization (hereinafter “optimization-based” methods) and purely analytical methods based on known quadrature rules (hereinafter “quadrature-based” methods). Note that methods based on discrete diffusive representations are also known as “non-classical” methods [12,4].

In [20], which deals with a fractional monodimensional wave equation, the fractional integral is split into two parts, namely a local and a historical one: while the former is approximated ad hoc, a Gauss–Legendre quadrature rule is employed

* Corresponding author.
E-mail addresses: florian.monteghetti@onera.fr (F. Monteghetti), denis.matignon@isae.fr (D. Matignon), estelle.piot@onera.fr (E. Piot).

https://doi.org/10.1016/j.apnum.2018.12.003

for the later, see [25] for an analysis. Another approach consists in directly using a quadrature rule, without any split. To get back to a finite interval, one can either truncate the semi-infinite integration domain [2] or use a change of variable [46,12,4].

In [2], Gauss–Legendre and Curtis–Clenshaw quadrature rules are used on a truncated domain. A method proposed in [46], based on a Gauss–Laguerre quadrature rule with a change of variable, has been widely investigated and led to the definitions of methods based instead on the Gauss–Jacobi quadrature rule [12,4], see [4] for a comparison that favors [4, Eq. (23)].

Optimization-based methods have also received scrutiny and enjoyed a wide range of applications, notably in wave propagation problems. A method based on a linear least squares optimization where the pole distribution is chosen a priori has been introduced in [16] for the identification of a lead acid battery impedance model using time-domain measurements. Further refinements have been proposed in [21], with application to a wide range of diffusive operators, and in [26], where a nonlinear least squares is compared with the method proposed in [4], mentioned above.

The objective of this paper is to investigate the discretization of diffusive representations using Gaussian quadrature, for application in the numerical solution of fractional differential equations as well as related eigenvalue problems. Inspired by classical works on numerical integration [10,1], a family of discretization methods that rely on the Gauss–Legendre quadrature rule is introduced and analyzed both theoretically and numerically. The analysis enables to pin down the most suitable method for applications. In particular, it emphasizes that the method must be tailored to the kernel at hand, by contrast with a one-size-fits-all approach. Numerical comparisons with existing discretization methods, both optimization and quadrature based, shed light on the practical interest of the proposed method. Additionally, it is shown on a numerical example that quadrature-based discretization methods are spectrally correct, i.e. that they yield an unpolluted and convergent approximation of the essential spectrum (linked to the fractional derivative), by contrast with optimization-based methods.

This paper is organized as follows. Section 2 recalls elementary facts about diffusive representations and introduces the proposed $Q_{\beta,N}$ discretization method, where β is a scalar parameter to be suitably chosen and N is the number of quadrature nodes. Section 3 presents an analysis of the method in the case of fractional operators, which highlights the dependency of β upon the order of the fractional operator. Numerical applications and comparisons are gathered in Section 4, where the $Q_{\beta,N}$ method is compared against two existing methods, one optimization-based and one quadrature-based. Section 5 investigates the use of a nonlinear least squares minimization to refine the poles and weights given by the $Q_{\beta,N}$ method.

2. Definition of the proposed quadrature-based discretization method

The purpose of this section is to introduce the proposed $Q_{\beta,N}$ discretization method, where β is a scalar parameter to be suitably chosen and N is the number of quadrature nodes. After some background on diffusive representations in Section 2.1, the method is defined in Section 2.2, namely in Definition 4.

2.1. Diffusive representation

In this paper, we consider the discretization of so-called *diffusive kernels*, expressed as

$$h(t) := \int_0^\infty e^{-\xi t} H(t)\, \mu(\xi)\, \mathrm{d}\xi \quad (t \in \mathbb{R}), \tag{1}$$

where H is the Heaviside or unit step function ($H(t) = 1$ for $t > 0$, null elsewhere) and $\mu \in \mathcal{C}((0,\infty))$ is the *diffusive weight*. By definition, diffusive kernels are locally integrable on $[0,\infty)$, i.e. $h \in L^1_{\mathrm{loc}}([0,\infty))$, so that the diffusive weight satisfies

$$\int_0^\infty \frac{\mu(\xi)}{1+\xi}\, \mathrm{d}\xi < \infty.$$

Note that, in general, h is not integrable over $(0,\infty)$. This class of kernels is physically linked to non-propagating diffusion phenomena, encountered in viscoelasticity [11,44,29], electromagnetics [18], and acoustics [38] [35, Chap. 2]. See [39,21,6, 26] and references therein for further background on diffusive representations and their applications. By defining the Laplace transform as

$$\hat{h}(s) := \int_0^\infty h(t) e^{-st}\, \mathrm{d}s \quad (\Re(s) > 0),$$

the identity (1) reads

$$\hat{h}(s) = \int_0^\infty \frac{\mu(\xi)}{s+\xi}\, \mathrm{d}\xi.$$

Remark 1. As defined herein, a diffusive kernel is a locally integrable completely monotone kernel on $(0,\infty)$. A diffusive kernel h is integrable on $(0,\infty)$ if and only if [19, Thm. 5.2.5]

$$\int_0^\infty \frac{\mu(\xi)}{\xi}\mathrm{d}\xi < \infty,$$

which is not the case for the kernels considered in this paper, see the three examples below.

Remark 2 *(Terminology).* In this paper, we use the following terminology: the *diffusive representation* of h is the identity (1), while the function μ is called the *diffusive weight*. This slightly differs from [39] where μ is called the diffusive representation of h. The quantity μ is also known under other names such as *spectral function* [18] or *relaxation spectrum* [29].

The computational interest of diffusive kernels is that, formally, the convolution operator $u \mapsto h \star u$ admits the following infinite-dimensional time-local realization

$$\begin{cases} \partial_t \varphi(t,\xi) = -\xi\varphi(t,\xi) + u(t),\ \varphi(0,\xi) = 0 \quad (\xi \in (0,\infty)), \\ h \star u(t) = \displaystyle\int_0^\infty \varphi(t,\xi)\mu(\xi)\,\mathrm{d}\xi, \end{cases} \tag{2}$$

where u is a causal input. A functional framework for this realization has been proposed in [11,44]. Let us now list three examples of diffusive operators covered by the discretization method introduced in Section 2.2.

1. The Riemann–Liouville fractional integral, defined as [41, § 2.3] [32]

   $$I^\alpha u := Y_\alpha \star u,$$

   where

   $$\alpha \in (0,1)$$

   and the fractional kernel is

   $$Y_\alpha(t) := \frac{H(t)}{\Gamma(\alpha)\,t^{1-\alpha}},\ \hat{Y}_\alpha(s) = \frac{1}{s^\alpha}. \tag{3}$$

   The associated diffusive weight is

   $$\mu_\alpha(\xi) := \frac{\sin(\alpha\pi)}{\pi\xi^\alpha}. \tag{4}$$

2. Another diffusive kernel is the zeroth-order Bessel function of the first kind [31, § 3.3]

   $$\begin{aligned} J_0(t)H(t) &= +e^{it}\int_0^\infty \frac{\mu_{1/2}(\xi)}{\sqrt{-\xi+2i}}e^{-\xi t}\,\mathrm{d}\xi + e^{-it}\int_0^\infty \frac{\mu_{1/2}(\xi)}{\sqrt{-\xi-2i}}e^{-\xi t}\,\mathrm{d}\xi \\ &= +2\Re\left[e^{it}\int_0^\infty \frac{\mu_{1/2}(\xi)}{\sqrt{-\xi+2i}}e^{-\xi t}\,\mathrm{d}\xi\right], \end{aligned} \tag{5}$$

   where $\mu_{1/2}$ is given by (4) and i is the unit imaginary number.

3. The fractional Caputo derivative, defined as [5] [40, § 2.4.1] [32]

   $$\mathrm{d}^\alpha u := I^{1-\alpha}\dot{u}, \tag{6}$$

   where $\dot{u}$ is the strong derivative. It formally admits the infinite-dimensional time-domain realization (contrast with (2))

   $$\begin{cases} \partial_t \varphi(t,\xi) = -\xi\varphi(t,\xi) + u(t),\ \varphi(0,\xi) = \dfrac{u(0)}{\xi} \quad (\xi \in (0,\infty)), \\ \mathrm{d}^\alpha u(t) = \displaystyle\int_0^\infty (-\xi\varphi(t,\xi) + u(t))\,\mu(\xi)\,\mathrm{d}\xi, \end{cases} \tag{7}$$

   where u is a sufficiently regular causal input. If $u(0) = 0$, then $\mathrm{d}^\alpha u$ matches the Riemann–Liouville fractional derivative.

These three convolution operators can be discretized using the $Q_{\beta,N}$ method, introduced in Section 2.2 below.

2.2. Discretization method

The causal kernel h given by (1) is discretized using N first-order kernels as

$$h(t) \simeq h_{\mathrm{num}}(t) := \sum_{n=1}^{N} \mu_n e^{-\xi_n t} H(t) \quad (t \in \mathbb{R}). \tag{8}$$

In the Laplace domain, this reads

$$\hat{h}(s) \simeq \hat{h}_{\mathrm{num}}(s) = \sum_{n=1}^{N} \frac{\mu_n}{s + \xi_n} \quad (\Re(s) > 0).$$

In this work, we seek to find an expression for (ξ_n, μ_n) that applies at least to the kernels listed in Section 2.1, whose diffusive weights μ are monotone on $(0, \infty)$ with a singularity at $\xi = 0$, which leads to the following assumption.

Assumption 3. *The diffusive weight $\mu \in \mathcal{C}((0,\infty))$ has a power-law singularity at $\xi = 0$, i.e.*

$$\mu(\xi) = \mathcal{O}\left(\frac{1}{\xi^{\alpha}}\right), \tag{9}$$

with $\alpha \in (0, 1)$.

Following classical works on numerical quadrature [10, Chap. 3] [1, § 5.6], the following two methods could be envisaged to deal with a singular integral like (1).

1. Consider μ as a weight function and define either a new set of Gauss nodes (if possible) or a new product quadrature rule with equidistant nodes [1, § 5.6].
2. Recover a continuous integrand using a change of variables. For example, for this integral, MATLAB® `integral` function uses the change of variable $\xi = \left(\frac{v}{1-v}\right)^2$, see [43, § 4.2].

To simplify the implementation, we choose the second method, i.e. we seek a suitable change of variables

$$\Psi : (-1, 1) \to (0, \infty),\ \Psi(-1) = 0,\ \Psi(1) = \infty,$$

so that the right-hand side of the identity

$$h(t) = \int_{-1}^{1} \mu\left(\Psi(v)\right) e^{-\Psi(v)t}\, \dot{\Psi}(v)\, \mathrm{d}v \tag{10}$$

can be accurately discretized using the Gauss–Legendre quadrature rule (v_n, w_n), thus yielding

$$\xi_n := \Psi(v_n),\ \mu_n := w_n \dot{\Psi}(v_n)\, \mu(\xi_n), \tag{11}$$

where $\dot{\Psi}$ denotes the derivative of Ψ. Given the singularity condition (9), a natural choice is [10, § 3.1] [1, § 5.6]

$$\Psi_\beta(v) := \left(\frac{1+v}{1-v}\right)^{\frac{1}{\beta}},\ \beta > 0. \tag{12}$$

This change of variables results from the composition of $v \mapsto \frac{1+v}{1-v}$, which maps $(-1, 1)$ to $(0, \infty)$, and the power law $v \mapsto v^{\frac{1}{\beta}}$. Using Ψ_β, the representation (10) reads

$$h(t) = \frac{2}{\beta} \int_{-1}^{1} e^{-t\left(\frac{1+v}{1-v}\right)^{\frac{1}{\beta}}} (1-v)^{-1-\frac{1}{\beta}} (1+v)^{\frac{1}{\beta}-1}\, \mu\left(\left(\frac{1+v}{1-v}\right)^{\frac{1}{\beta}}\right) \mathrm{d}v, \tag{13}$$

which leads to the definition of the $Q_{\beta,N}$ discretization method given below.

Definition 4. The $Q_{\beta,N}$ discretization of (1) is (8) with

$$\xi_n := \left(\frac{1+v_n}{1-v_n}\right)^{\frac{1}{\beta}},\ \mu_n := w_n \frac{2}{\beta} (1+v_n)^{\frac{1}{\beta}-1} (1-v_n)^{-1-\frac{1}{\beta}}\, \mu\left(\xi_n\right), \tag{14}$$

where (v_n, w_n) is the Gauss–Legendre quadrature rule [1, § 5.3].

Intuitively, one may expect the best value for β to be dependent on properties of the diffusive weight μ, such as the value of α in (9). Section 3 investigates this for the case of fractional operators.

3. Analysis for fractional operators

The purpose of this section is to show that, for the fractional kernel (3), the best practical value for β is given by (22). The theoretical analysis is presented in Section 3.1 and examples of approximation errors are provided in Section 3.2.

3.1. Theoretical analysis

Let us recall the following standard theorem.

Theorem 5 *(Convergence rate). Let (v_n, w_n) be the Gauss–Legendre quadrature rule and p a nonnegative integer. If $f \in \mathcal{C}^p([-1,1])$, then*

$$\lim_{N\to\infty} N^p \left| \int_{-1}^{1} f(v)\,dv - \sum_{n=1}^{N} w_n f(v_n) \right| = 0.$$

In particular, if $f \in \mathcal{C}^\infty([-1,1])$ then spectral convergence is achieved.

Proof. Since f is at least continuous on $[-1,1]$, we have the estimate [1, Thm. 5.4]

$$\left| \int_{-1}^{1} f(v)\,\mathrm{d}v - \sum_{n=1}^{N} w_n f(v_n) \right| \le 4 \inf_{\deg q \le 2N-1} \|f - q\|_{L^\infty([-1,1])}.$$

The conclusion follows from a polynomial approximation result [13, Thm. I.VIII].

Given the above result, to find the optimal value for β in the $Q_{\beta,N}$ discretization, it is sufficient to study the regularity of the integrand in (13). Let us now focus on the fractional kernel (3), which is the simplest kernel that satisfies (9). A first convergence result is summarized in the proposition below. □

Proposition 6. *Let $\beta > 0$, $N \in \mathbb{N}^*$, and $Y_{\alpha,\mathrm{num}}$ be the $Q_{\beta,N}$ discretization of Y_α with $\alpha \in (0,1)$. If*

$$\beta \le 1 - \alpha, \tag{15}$$

then $Y_{\alpha,\mathrm{num}}(t) \to Y_\alpha(t)$ as $N \to \infty$ for any $t > 0$. If, additionally,

$$\frac{1}{\beta} \in \mathbb{N}, \quad \frac{\alpha}{\beta} \in \mathbb{N}, \tag{16}$$

then $\left|Y_\alpha(t) - Y_{\alpha,\mathrm{num}}(t)\right| \underset{k\to\infty}{=} \mathcal{O}\left(n^{-k}\right)$ for every positive integer k and $t > 0$.

Proof. The diffusive representation (13) of the fractional kernel (3) reads

$$Y_\alpha(t) = \int_{-1}^{1} \Phi_\beta(t,v)\,\mathrm{d}v, \tag{17}$$

with

$$\Phi_\beta(t,v) := \frac{2\sin(\alpha\pi)}{\pi\beta} e^{-t\left(\frac{1+v}{1-v}\right)^{\frac{1}{\beta}}} (1-v)^{-1+\frac{\alpha-1}{\beta}} (1+v)^{-1+\frac{1-\alpha}{\beta}}.$$

Since $\Phi_\beta(t,\cdot) \in \mathcal{C}^\infty((-1,1))$, the only task is to investigate the singularities at -1 and 1. There is no singularity at $v = 1$ as long as $t > 0$, since $x \mapsto \frac{e^{-t\left(1+\frac{2}{x}\right)^{\frac{1}{\beta}}}}{x^{1+\frac{1-\alpha}{\beta}}}$ is infinitely differentiable at 0^+, without assumption on α and β. Since $v \mapsto e^{-t\left(\frac{1+v}{1-v}\right)^{\frac{1}{\beta}}}$ has a limit as $v \to -1^+$, the integrand $\Phi_\beta(t,\cdot)$ is continuous if and only if (15) holds. Furthermore, $\Phi_\beta(t,\cdot) \in \mathcal{C}^\infty([-1,1])$ if and only if (15) and (16) hold. The conclusion then follows from Theorem 5. □

From Proposition 6, a convergence result on $\|Y_\alpha - \tilde{Y}_\alpha\|_{L^1(\epsilon,T)}$ for any $\epsilon > 0$ and $T > \epsilon$ can be readily deduced, although this is not sufficient for time-domain computations. Indeed, for an input $u \in L^2(0,T) \cap L^\infty(0,T)$, we have the straightforward estimate

$$\frac{|h \star u(T)|}{\|u\|_{L^\infty(0,T)}} \leq \|h\|_{L^1(0,T)}, \tag{18}$$

which justifies an interest in approximating the L^1 norm of h. This requires an additional constraint on β, see Proposition 7.

Proposition 7. *Let $\beta > 0$, $N \in \mathbb{N}^*$, and $Y_{\alpha,\mathrm{num}}$ be the $Q_{\beta,N}$ discretization of Y_α with $\alpha \in (0,1)$. If both (15) and*

$$\beta \leq \alpha \tag{19}$$

hold, then $\lim_{N\to\infty} \|Y_{\alpha,\mathrm{num}}\|_{L^1(0,T)} = \|Y_\alpha\|_{L^1(0,T)}$ for any $T > 0$. If, additionally, (16) holds, then this convergence is spectral.

Proof. Let $T > 0$. Since $\alpha \in (0,1)$, we have $\Phi_\beta \in L^1((0,T) \times (-1,1))$. The Fubini theorem yields

$$\|Y_\alpha\|_{L^1(0,T)} = \int_{-1}^{1} \Theta_\beta(T,v)\,\mathrm{d}v, \tag{20}$$

where

$$\Theta_\beta(T,v) = \frac{2\sin(\alpha\pi)}{\pi\beta}\left[1 - e^{-T\left(\frac{1+v}{1-v}\right)^{\frac{1}{\beta}}}\right](1-v)^{-1+\frac{\alpha}{\beta}}(1+v)^{-1-\frac{\alpha}{\beta}}.$$

It is sufficient to investigate the regularity of $\Theta_\beta(T,\cdot)$ at $v = \pm 1$. $\Theta_\beta(T,\cdot)$ is continuous at 1 if and only if (19) holds. Expanding around $v = -1$ yields

$$\left[1 - e^{-t\left(\frac{1+v}{1-v}\right)^{\frac{1}{\beta}}}\right](1+v)^{-1-\frac{\alpha}{\beta}} = t(1+v)^{\frac{1-\alpha}{\beta}-1}(1-v)^{-\frac{1}{\beta}} + \cdots,$$

hence continuity at -1 is achieved if and only if (15) holds. If, additionally, (16) is assumed, then $\Theta_\beta(T,\cdot) \in \mathcal{C}^\infty([-1,1])$. □

Remark 8. Proposition 7 states convergence *of* the L^1 norm, but not convergence *in* the L^1 norm, i.e. $\lim_{N\to\infty} \|Y_\alpha - Y_{\alpha,\mathrm{num}}\|_{L^1(0,T)} = 0$ (which implies convergence of the L^1 norm). This proposition can therefore be deemed insufficient in view of the estimate (18); however, it gives a second constraint on β, namely (19), which is practically useful.

Remark 9 *(Frequency domain).* A similar study in the frequency domain reaches the same conclusion. For any $s \neq 0$ with $\Re(s) \geq 0$, we have

$$\hat{Y}_\alpha(s) = \frac{2}{\beta}\frac{\sin(\alpha\pi)}{\pi}\int_{-1}^{1} \frac{(1-v)^{\frac{\alpha}{\beta}-1}(1+v)^{\frac{1-\alpha}{\beta}-1}}{s(1-v)^{\frac{1}{\beta}} + (1+v)^{\frac{1}{\beta}}}\,\mathrm{d}v,$$

and the integrand is continuous on $[-1,1]$ if and only if (15, 19) hold. If, furthermore, (16) holds, then the integrand is infinitely smooth and we have spectral convergence for

$$\left|\hat{Y}_\alpha(s) - \sum_{n=1}^{N} \frac{\mu_n}{s+\xi_n}\right|.$$

For any $\omega_m > 0$, we readily deduce that $\|i\omega\hat{Y}_\alpha(i\omega) - i\omega\sum_{n=1}^{N}\frac{\mu_n}{i\omega+\xi_n}\|_{L^2(-\omega_m,\omega_m)}$ has the same convergence properties.

Practical choice of β

Based on the above results, the following rules can be followed to choose β in practice.

1. If $\alpha \in (0,1) \cap \mathbb{Q}$ such that $\alpha = \frac{n_0}{n_1}$ with $n_i \in \mathbb{N}^*$, then

$$\beta_1 := \frac{1}{n_1} \tag{21}$$

satisfies the condition (15), (16), and (19) so that the $Q_{\beta_1,N}$ method yields a spectrally convergent approximation. This value is also suited for $\alpha \in (0,1) \cap (\mathbb{R}\backslash\mathbb{Q})$ with $\alpha \simeq \frac{n_0}{n_1}$.

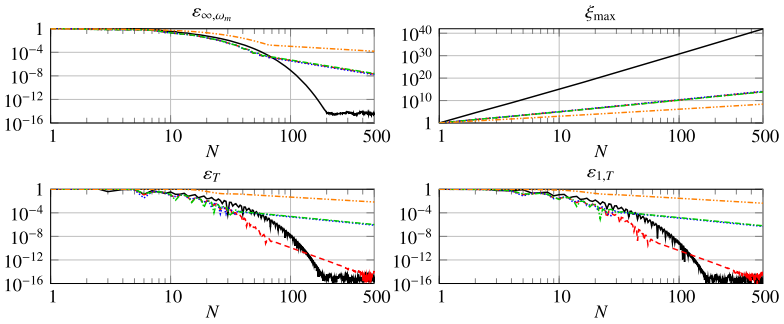

Fig. 1. Errors (23, 24, 25) and maximum pole (26) for $h = Y_\alpha$ with $\alpha = \frac{5}{8}$. (——) $Q_{\beta,N}$ with $\beta = \beta_1 = \frac{1}{8}$. (- - - -) $Q_{\beta,N}$ with $\beta = \beta_2$. (·····) $Q_{\beta,N}$ with $\beta = \beta_2 \times 0.99$. (-·-·-) $Q_{\beta,N}$ with $\beta = \beta_2 \times 1.01$. (-··-··-) $Q_{\beta,N}$ with $\beta = \beta_3$.

2. The conditions (15) and (19) suggest using a larger value of β, namely

$$\beta_2 := \min(\alpha, 1-\alpha), \tag{22}$$

which yields at least a convergent approximation from Propositions 6 and 7. Section 3.2 below shows that β_2 is the most interesting choice for moderate values of N.

3.2. Numerical illustrations

To investigate numerically the influence of β on the convergence of the $Q_{\beta,N}$ method, we define three errors. The first one is in the frequency domain

$$\varepsilon_{\infty,\omega_m} := \left\| 1 - \frac{\hat{h}_{\text{num}}}{\hat{h}}(i\omega) \right\|_{L^\infty(-\omega_m,\omega_m)}, \tag{23}$$

with $\omega_m > 0$ a given angular frequency. The second and third ones are in the time domain, namely

$$\varepsilon_T := \left| 1 - \frac{h_{\text{num}}}{h} \right|(T), \ \varepsilon_{1,T} := \frac{\left| \|h\|_{L^1(0,T)} - \|h_{\text{num}}\|_{L^1(0,T)} \right|}{\|h\|_{L^1(0,T)}}, \tag{24}$$

with $T > 0$. From now on, we set

$$\omega_m = T = 10^4, \tag{25}$$

so that we consider broadband approximations of the kernel h. We first consider $h = Y_\alpha$, covered by the results of Section 3.1, with four values of α and then conclude with $h = J_0$. Computations are done with double precision floating point.

The case $h = Y_\alpha$ with $\alpha = \frac{5}{8} \simeq 0.62 \geq \frac{1}{2}$ is shown in Fig. 1. The choice $\beta_1 = \frac{1}{8}$ achieves spectral convergence, with saturation at double precision, as expected from Section 3.1. The value $\beta_2 = 1 - \alpha$ does not converge spectrally, but it provides a better approximation for moderate values of N. The value

$$\beta_3 := \max(\alpha, 1-\alpha),$$

which does not satisfy (15), is the least interesting option. The sensitivity of the errors obtained with $\beta = \beta_2$ is highlighted by the curves corresponding to $\beta = 0.99 \times \beta_2$ and $\beta = 1.01 \times \beta_2$, which are significantly worse in the time domain. These error plots highlight that the time-domain norms do add information: here, the sensitivity to β_2 cannot be seen in the frequency domain for instance, while it is the opposite for other values of α covered below.

The upper right plot of Fig. 1 gives the maximum pole

$$\xi_{\max} := \max_n \xi_n. \tag{26}$$

This quantity is especially important when using an explicit scheme to advance the realization (2) in time, since the time step typically scales as $\mathcal{O}(\xi_{\max}^{-1})$. The plot shows that

$$\xi_{\max} = \mathcal{O}(N^{\frac{2}{\beta}}).$$

Given that the higher $\xi_{\max}$, the more costly the time integration, one may expect that a higher value of $\xi_{\max}$ systematically yields a more accurate discretization. However, this need not be the case: although this is indeed verified for β_1, β_2, and β_3, the values $0.99 \times \beta_2$ and $1.01 \times \beta_2$ have a value of $\xi_{\max}$ similar to $\beta = \beta_2$, but give significantly worse approximations.

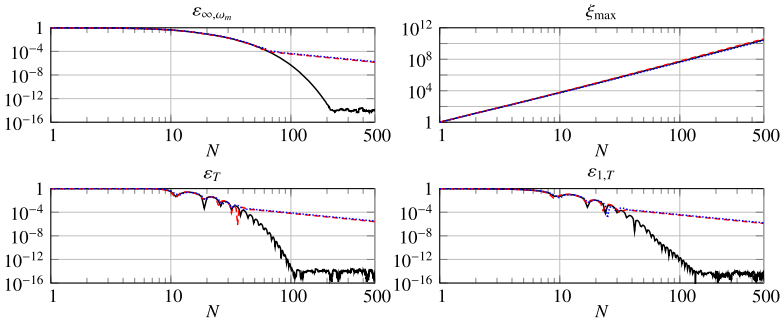

Fig. 2. Errors (23, 24, 25) and maximum pole (26) for $h = Y_\alpha$ with $\alpha = \frac{1}{2}$. (——) $Q_{\beta,N}$ with $\beta = \frac{1}{2}$. (- - - -) $Q_{\beta,N}$ with $\beta = \beta_2 \times 0.99$. (·····) $Q_{\beta,N}$ with $\beta = \beta_2 \times 1.01$.

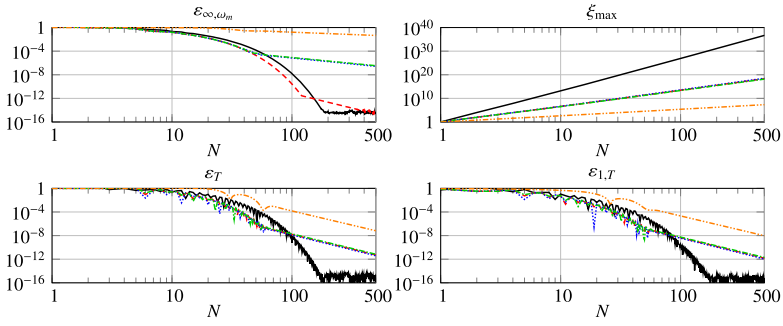

Fig. 3. Errors (23, 24, 25) and maximum pole (26) for $h = Y_\alpha$ with $\alpha = \frac{2}{7}$. (——) $Q_{\beta,N}$ with $\beta = \beta_1 = \frac{1}{7}$. (- - - -) $Q_{\beta,N}$ with $\beta = \beta_2$. (·····) $Q_{\beta,N}$ with $\beta = \beta_2 \times 0.99$. (-·-·-) $Q_{\beta,N}$ with $\beta = \beta_2 \times 1.01$. (-··-··-) $Q_{\beta,N}$ with $\beta = \beta_3$.

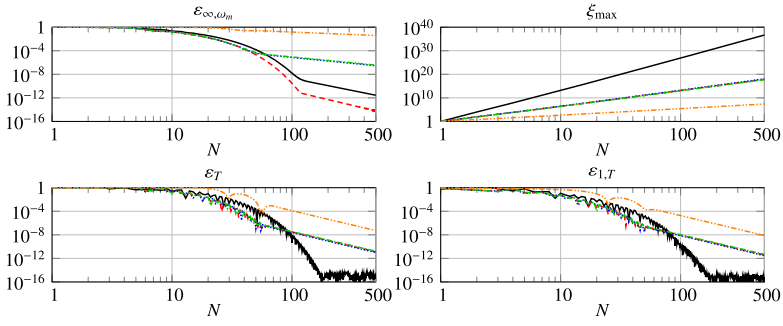

Fig. 4. Errors (23, 24, 25) and maximum pole (26) for $h = Y_\alpha$ with $\alpha = \frac{\sqrt{2}-1}{\sqrt{2}}$. (——) $Q_{\beta,N}$ with $\beta = \beta_1 = \frac{1}{7}$. (- - - -) $Q_{\beta,N}$ with $\beta = \beta_2$. (·····) $Q_{\beta,N}$ with $\beta = \beta_2 \times 0.99$. (-·-·-) $Q_{\beta,N}$ with $\beta = \beta_2 \times 1.01$. (-··-··-) $Q_{\beta,N}$ with $\beta = \beta_3$.

Fig. 2 plots the error graphs for $\alpha = \frac{1}{2}$. Here, the values β_1, β_2, and β_3 are identical so that the $Q_{\beta,N}$ discretization enjoys spectral convergence, with double precision on ε_T reached for around 100 variables. The sensitivity to a change in β around β_2 can be seen in both frequency-domain and time-domain errors: although the values of $\xi_{\max}$ remain close, the approximations are significantly worse for $0.99 \times \beta_2$ and $1.01 \times \beta_2$.

The conclusions for $\alpha = \frac{2}{7} \simeq 0.28 \leq \frac{1}{2}$, shown in Fig. 3, are identical to $\alpha = \frac{5}{8}$. The only difference is that the sensitivity to β_2 is only seen in the frequency-domain norm $\varepsilon_{\infty,\omega_m}$. Fig. 4 shows the errors obtained for $\alpha = \frac{\sqrt{2}-1}{\sqrt{2}} \simeq 0.29$, a value close to $\frac{2}{7}$ but irrational. The main difference is the error obtained for $\beta = \beta_1$, which is less accurate in the frequency domain compared to $\alpha = \frac{2}{7}$. However, the hierarchy between the $Q_{\beta,N}$ methods is identical, and the other errors are similar. The choices $\beta = \frac{1}{41}$ (justified by $\alpha \simeq \frac{12}{41}$) and $\beta = \frac{1}{3}$ (justified by $\alpha \simeq \frac{1}{3}$), not shown here, deliver poorer results. Overall, Fig. 4 illustrates that the irrationality of α is not a major concern in practice.

In summary, Figs. 1–4 show that, for moderate values of N, the $Q_{\beta,N}$ method with $\beta = \beta_2$ delivers satisfactory convergence results for any $\alpha \in (0, 1)$, rational or irrational. In addition, the fact that $\beta_2 \geq \beta_1$ implies that the $Q_{\beta_2,N}$ method yields a lower maximum pole (26) than $Q_{\beta_1,N}$, which is of particular interest for time-domain simulations. These two properties implies that the choice $\beta = \beta_2$ is satisfactory in practice.

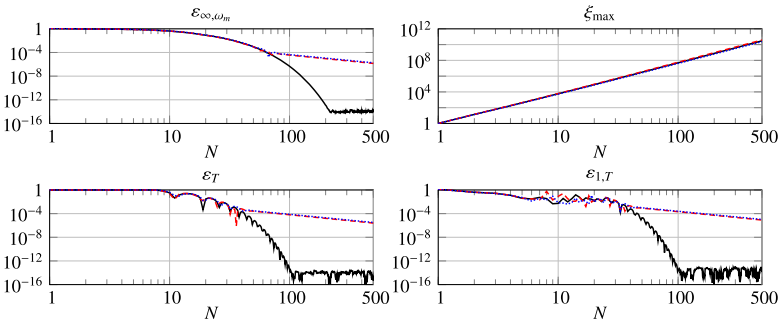

Fig. 5. Errors (23, 24, 25) and maximum pole (26) for $h = J_0$. (——) $Q_{\beta,N}$ with $\beta = \frac{1}{2}$. (- - - -) $Q_{\beta,N}$ with $\beta = \beta_2 \times 0.99$. (·····) $Q_{\beta,N}$ with $\beta = \beta_2 \times 1.01$.

Fig. 5 gives the errors obtained in approximating the Bessel function J_0, whose diffusive representation is given by (5): the results are similar to that shown in Fig. 2 for the fractional kernel of order $1/2$, since the diffusive weights of both kernels have a similar behavior.

The computational merits of the $Q_{\beta,N}$-method with $\beta = \beta_2$ are further investigated in Section 4, where numerical applications are gathered.

4. Numerical applications and comparisons

The purpose of this section is to investigate the computational properties of the $Q_{\beta,N}$ method as well as compare them to those of two existing methods: one quadrature-based, recalled in Section 4.1, and one optimization-based, recalled in Section 4.2. The comparison is carried out in the other three sections: Section 4.3 gathers approximation errors, Section 4.4 focuses on the simulation of a fractional differential equation, and Section 4.5 investigates spectral correctness, which turns out to be an important feature of the $Q_{\beta,N}$ method, and, more generally, of quadrature-based methods.

4.1. Birk–Song quadrature method

After reviewing existing methods, notably [46] and [12], Birk and Song proposed the change of variable $\xi = \Psi_\beta(v)$ with $\beta = \frac{1}{4}$. However, they propose to use a Gauss–Jacobi quadrature rule instead of a Gauss–Legendre one (thereby introducing a singularity at $v = -1$ in the integrand), which leads to the discrete representation [4, Eq. (23)]

$$\xi_n := \left(\frac{1-\tilde{v}_n}{1+\tilde{v}_n}\right)^4, \quad \mu_n := 8\frac{\sin(\alpha\pi)}{\pi}\frac{\tilde{w}_n}{(1+\tilde{v}_n)^4}, \tag{27}$$

where $(\tilde{v}_n, \tilde{w}_n)$ is the Gauss–Jacobi quadrature rule for the weight function $v \mapsto (1-v)^{2\overline{\alpha}+1}(1+v)^{-(2\overline{\alpha}-1)}$ with $\overline{\alpha} := 1-2\alpha$. (Beware that, in [4, Eq. (23)], "α" denotes the order of the Caputo derivative, whereas herein, α is the order of the fractional integral.)

4.2. Optimization method

We briefly recall here the optimization method defined in [21, § 4.3], which consists in a least squares optimization. The main challenge of such an optimization is that $\hat{h}_{\text{num}}$ is nonlinear with respect to the poles $(\xi_n)_n$, which furthermore have a wide variation since theoretically $\xi \in (0, \infty)$. To avoid this computational difficulty the method proceeds as follows.

1. The three input parameters, namely $N \in [\![2, \infty[\![$, $\xi_{\min} > 0$, and $\xi_{\max} > \xi_{\min}$ are *chosen*.
2. The N poles ξ_n are logarithmically spaced in $[\xi_{\min}, \xi_{\max}]$:
$$\xi_n = \xi_{\min}\left(\frac{\xi_{\max}}{\xi_{\min}}\right)^{\frac{n-1}{N-1}} \quad (n \in [\![1, N]\!]).$$
3. Let $A := \left[(i\omega_k + \xi_n)^{-1}\right]_{k,n} \in \mathbb{C}^{K\times N}$ and $b := \left[\hat{h}(i\omega_k)\right]_k \in \mathbb{C}^K$, where the K angular frequencies ω_k are also logarithmically spaced in $[\xi_{\min}, \xi_{\max}]$. The N weights μ_n are computed with a linear least squares minimization of
$$J(\mu) := \|C\mu - d\|_2^2 = \sum_{k=1}^{K}\left|\sum_{n=1}^{N}\frac{\mu_n}{i\omega_k + \xi_n} - \hat{h}(i\omega_k)\right|^2, \tag{28}$$
where C and d are given by

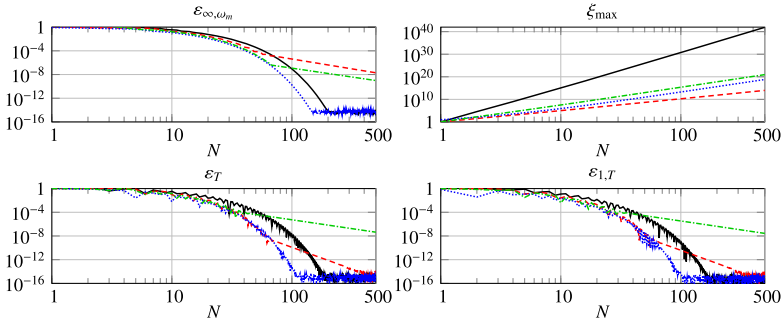

Fig. 6. Errors (23, 24, 25) and maximum pole (26) for $h = Y_\alpha$ with $\alpha = \frac{5}{8}$. (——) $Q_{\beta,N}$ with $\beta = \beta_1 = \frac{1}{8}$. (- - - -) $Q_{\beta,N}$ with $\beta = \beta_2$. (·····) Birk–Song method (27). (-·-·-) $Q_{\beta,N}$ with $\beta = \frac{1}{4}$.

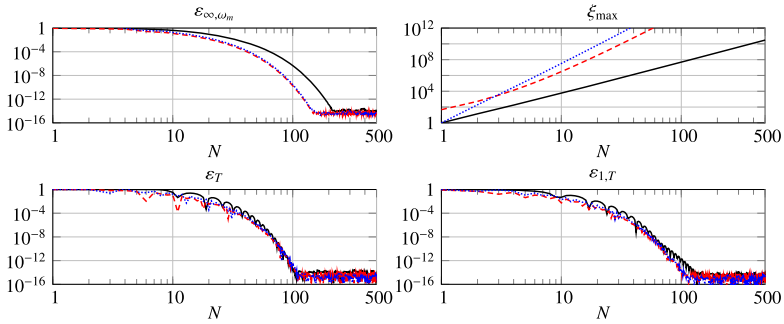

Fig. 7. Errors (23, 24, 25) and maximum pole (26) for $h = Y_\alpha$ with $\alpha = \frac{1}{2}$. (——) $Q_{\beta,N}$ with $\beta = \frac{1}{2}$. (- - - -) Birk–Song method (27). (·····) $Q_{\beta,N}$ with $\beta = \frac{1}{4}$.

$$C := \begin{bmatrix} \Re(A) \\ \Im(A) \end{bmatrix} \in \mathbb{R}^{2K\times N},\ d := \begin{bmatrix} \Re(b) \\ \Im(b) \end{bmatrix} \in \mathbb{R}^{2K}.$$

Provided that $2K > N$ the problem is overdetermined and can be directly solved by a pseudo-inverse. The reality of the weights μ_n is enforced through the definition of C and d, which separates real and imaginary parts. However, note that the sign of each μ_n is unconstrained.

This technique is particularly suited for time-domain simulations, where $\xi_{\max}$ is naturally known (from e.g. the minimum acceptable time step or the maximum frequency of interest in wave propagation problems); it can also handle more complex representations that involve additional poles. For a given N and $\xi_{\max}$, there is usually an optimal range for the lower bound $\xi_{\min}$, which governs the long-time behavior of h_{num}, which must not be chosen too small. For the diffusive kernels considered herein, a logarithmic spacing of the poles ξ_n is satisfactory (a linear spacing yields poorer results). In all the applications presented in this section, we set

$$K = 10^4. \tag{29}$$

Remark 10. There is an inherent difficulty when comparing the above optimization method with the $Q_{\beta,N}$ method, since both do not have the same number of parameters: 1 for the $Q_{\beta,N}$ method (namely the number of quadrature nodes N, since β has been chosen to be β_2 following the analysis of Section 3), 3 for the optimization method (namely N and the minimum and maximum poles $\xi_{\min}$ and $\xi_{\max}$). In all the results presented below, the parameters $\xi_{\min}$ and $\xi_{\max}$ of the optimization method have been empirically chosen to yield the best results.

4.3. Approximation errors

In the spirit of Section 3.2, here are gathered comparisons of the approximation errors.

Comparison with the Birk–Song method. A comparison between the $Q_{\beta,N}$ method and the Birk–Song method (27) is shown in Fig. 6 for $\alpha = \frac{5}{8}$ and Fig. 7 for $\alpha = \frac{1}{2}$.

Let us first consider the case $\alpha = \frac{5}{8}$. The behavior of $\varepsilon_{\infty,\omega_m}$ highlights the accuracy of the Birk–Song method in the frequency domain, where it outperforms the three $Q_{\beta,N}$ methods, the closest being the one corresponding to $\beta = \frac{1}{4}$ for

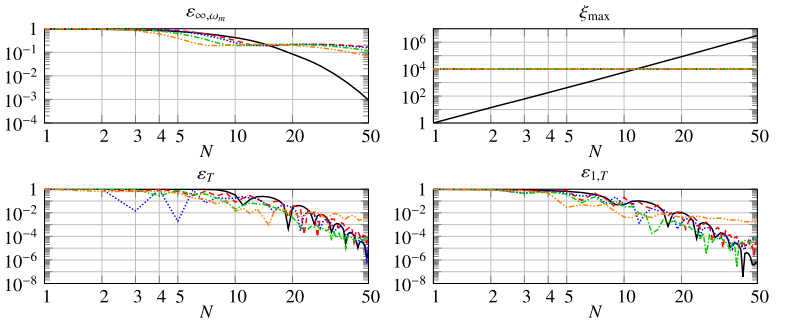

Fig. 8. Errors (23, 24, 25) and maximum pole (26) for $h = Y_\alpha$ with $\alpha = \frac{1}{2}$. (——) $Q_{\beta,N}$ with $\beta = \beta_2$. Optimization with $\xi_{\max} = 10^4$: (----) $\xi_{\min} = 10^{-16}$, (·····) $\xi_{\min} = 10^{-14}$, (-·-·-) $\xi_{\min} = 10^{-10}$, (-···-) $\xi_{\min} = 10^{-6}$.

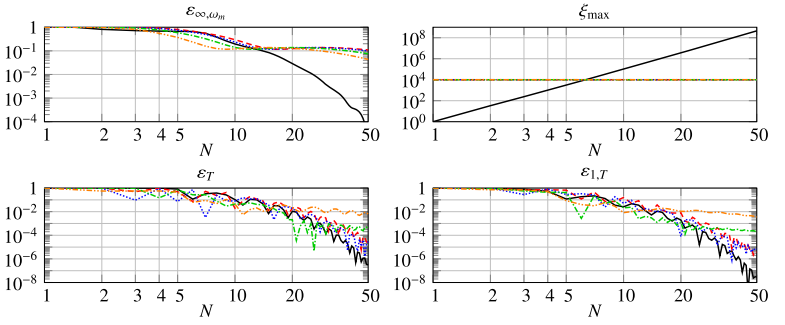

Fig. 9. Errors (23, 24, 25) and maximum pole (26) for $h = Y_\alpha$ with $\alpha = \frac{5}{8}$. (——) $Q_{\beta,N}$ with $\beta = \beta_2$. Optimization with $\xi_{\max} = 10^4$: (----) $\xi_{\min} = 10^{-16}$, (·····) $\xi_{\min} = 10^{-14}$, (-·-·-) $\xi_{\min} = 10^{-10}$, (-···-) $\xi_{\min} = 10^{-6}$.

$N \leq 70$, as one may expect from the change of variable that defines the Birk–Song method. A similar trend is seen in the time domain, although there the closest $Q_{\beta,N}$ method for $N \leq 70$ is that obtained with $\beta = \beta_2$. All the methods are closer for $\alpha = \frac{1}{2}$, at least in the time domain, and the method $Q_{\beta,N}$ with $\beta = \frac{1}{4}$ has almost identical convergence properties to that of the Birk–Song method.

As already mentioned when comparing the various $Q_{\beta,N}$ methods in Section 3.2, the graphs of $\varepsilon_{\infty,\omega_m}$, ε_T, and $\varepsilon_{1,T}$ alone are not sufficient to compare discretization methods: one must take into account the value of $\xi_{\max}$, shown in the top right plot of Figs. 6 and 7. These plots show that for the Birk–Song method we have $\xi_{\max} = \mathcal{O}(N^{\frac{2}{\beta}})$ with $\beta = \frac{1}{4}$, i.e. $\xi_{\max} = \mathcal{O}(N^8)$, which implies significantly larger values that the $Q_{\beta_2,N}$ method recommended from the analysis of Section 3.2. The impact of these large values of $\xi_{\max}$ is of concern when using explicit time-marching scheme, see Section 4.4.

Comparison with the optimization method. The errors for $\alpha = \frac{1}{2}$ and $\alpha = \frac{5}{8}$ are plotted in Figs. 8 and 9, respectively. Given the results of Section 3, only the $Q_{\beta,N}$ method with $\beta = \beta_2$ is considered. For the (three-parameter) optimization method, we choose $\xi_{\max} = \omega_m = 10^4$ and plot the errors for various values of $\xi_{\min}$: the result shows that the optimal value of $\xi_{\min}$ does strongly depend upon N, so that $\xi_{\min}$ is not straightforward to choose a priori. However, provided that the value of $\xi_{\min}$ is well-chosen, the optimization method can outperform the $Q_{\beta_2,N}$ method on a range of N, which justifies its popularity in large-scale applications where the value of N is critical. Note that, by contrast with Section 3.2, the comparison is restricted to $N \in [\![1, 50]\!]$, since outside of this interval, the maximum pole $\xi_{\max}$ of the $Q_{\beta_2,N}$ method is significantly larger than 10^4 so that the comparison would not be fair. In summary, here, the main advantage of the $Q_{\beta_2,N}$-method is that it has just one parameter.

To refine the computed poles and weights, one may consider the use of a nonlinear least squares minimization by adding the following fourth step to the three-step optimization method described in Section 4.2:

4. Compute N new weights $(\mu_n)_n$ and poles $(\xi_n)_n$ with a nonlinear least squares minimization of the right-hand side of (28), starting from the poles chosen in step 2 and the weights obtained in step 3, with the following linear constraints

$$\mu_n \geq 0, \ \xi_n \geq 0, \ \xi_n \leq \xi_{\max} \quad (n \in \mathbb{N}).$$

Fig. 10 shows the approximation errors obtained using the trust-region algorithm implemented in MATLAB® `lsqnonlin`. Both the convergence speed of the nonlinear optimization stage and the quality of the end result strongly depend upon the initial poles distribution, which makes this method unpractical (for example, the case $\xi_{\min} = 10^{-14}$ is difficult to

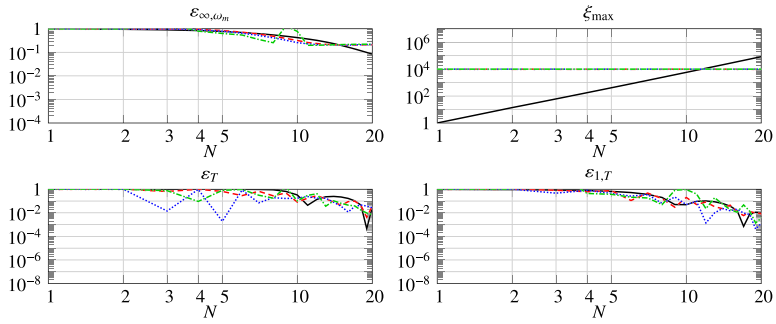

Fig. 10. Errors (23, 24, 25) and maximum pole (26) for $h = Y_\alpha$ with $\alpha = \frac{1}{2}$. (——) $Q_{\beta,N}$ with $\beta = \beta_2$. Four-stage optimization with $\xi_{\max} = 10^4$: (- - - -) $\xi_{\min} = 10^{-16}$, (·····) $\xi_{\min} = 10^{-14}$, (-·-·-) $\xi_{\min} = 10^{-10}$. (Same parameters as Fig. 8.)

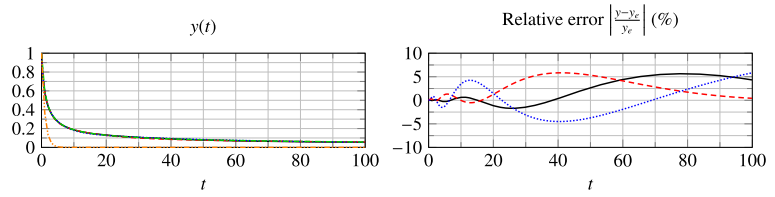

Fig. 11. FDE (30) for $y_0 = 1$ and $g = 1$. Numerical solutions computed with RKF84, $\Delta t = 9 \times 10^{-3}$, and $N = 6$. (——) $Q_{\beta,N}$ with $\beta = \beta_2$ ($\xi_{\min} = 1.221 \times 10^{-3}$, $\xi_{\max} = 8.189 \times 10^2$). (- - - -) Optimization ($\xi_{\min} = 10^{-3}$, $\xi_{\max} = 10^2$). (·····) Optimization ($\xi_{\min} = 10^{-4}$, $\xi_{\max} = 10^2$). (Left only) (-·-·-) Exact solution (31), (-··-··-) Exact solution for $g = 0$.

converge while $\xi_{\min} = 10^{-16}$ is almost instantaneous.). Furthermore, the approximation error graphs show that it is unsatisfactory, so that it is not worth considering in practice. In fact, nonlinear optimization is best used in *combination* with quadrature rules: this is investigated in Section 5. This nonlinear four-stage optimization method is not further considered in the remaining of this section: "optimization method" will denote the three-step method described in Section 4.2.

4.4. Fractional differential equation

Let us consider the following scalar fractional differential equation

$$\dot{y}(t) = a y(t) - g\, \mathrm{d}^{\frac{1}{2}} y(t), \;\; y(0) = y_0 \quad (t > 0), \tag{30}$$

where $\dot{y}$ is the strong derivative and $\mathrm{d}^{\frac{1}{2}}$ is the Caputo derivative defined in (6). The exact solution of (30) can be expressed using the Mittag-Leffler function $E_{\alpha,\beta}$ as [32, Ex. 1.6]

$$y_e(t) := \frac{y_0}{\lambda_1 - \lambda_2} \left[\lambda_1 E_{1/2,1}\left(\lambda_2 \sqrt{t}\right) - \lambda_2 E_{1/2,1}\left(\lambda_1 \sqrt{t}\right) \right], \tag{31}$$

where λ_1 and λ_2 are the roots of $s \mapsto s^2 + gs - a$. The left plot of Fig. 11 shows the exact solution on $[0, t_f]$ with $t_f = 100$, $y_0 = 1$, and for both $g = 0$ (i.e. standard ODE) and $g = 1$ to highlight the effect of the fractional derivative. To accurately evaluate y_e, we rely on the algorithm proposed in [17].

Comparison with the optimization method. We seek to compute numerical solutions of (30) with a relative accuracy of, say, 6%. With the $Q_{\beta,N}$ method, the sole parameter of which is N, this accuracy target is attained for any $N \geq 6$, so that we set $N = 6$ for the optimization method as well. The corresponding numerical solutions are shown in Fig. 11, which also plots the relative error for other values of $\xi_{\min}$ and $\xi_{\max}$. Time-integration is performed using a fourth-order eight-stage explicit Runge–Kutta method, namely the RKF84 from [45, Tab. A.9], with a timestep of $\Delta t = 9 \times 10^{-3}$, which is the maximum stable time step for all methods. As expected from Section 3.2, both methods yield similar results.

Comparison with the Birk–Song method. For the time step $\Delta t = 9 \times 10^{-3}$, used in Fig. 11, the Birk–Song method (27) yields a stable result only for $N \leq 2$. For instance, for $N = 3$, the stability timestep is found to be $\Delta t_{\max} = 2.37 \times 10^{-3}$, which is a significant reduction. This can be explained by the large values of $\xi_{\max}$, already highlighted in Section 4.3. This timestep reduction could be balanced by an accuracy increase. To investigate this, Fig. 12 plots a comparison between the Birk–Song and $Q_{\beta_{2,N}}$ methods at a timestep well-below the stability limit, namely $\Delta t = 10^{-3}$ for $N = 3$. On this example, the $Q_{\beta,N}$ method is more accurate, in addition to having a larger stability limit.

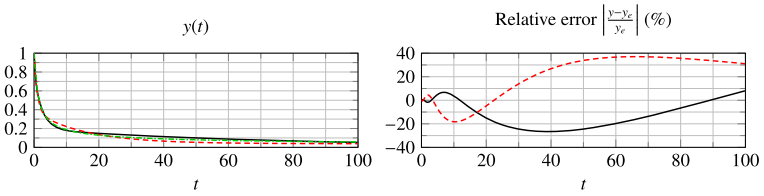

Fig. 12. FDE (30) for $y_0 = 1$ and $g = 1$. Numerical solutions computed with RKF84, $\Delta t = 10^{-3}$ and $N = 3$. (——) $Q_{\beta,N}$ with $\beta = \beta_2$ ($\xi_{\min} = 1.613 \times 10^{-2}$, $\xi_{\max} = 6.198 \times 10^{1}$). (- - - -) Birk–Song method (27) ($\xi_{\min} = 3.139 \times 10^{-4}$, $\xi_{\max} = 3.185 \times 10^{3}$). (Left only) (-·-·-) Exact solution (31).

4.5. Eigenvalue approach to stability

To conclude this section on numerical applications, let us consider a case where the $Q_{\beta,N}$ and optimization methods have radically different properties. We are interested in studying the stability of the solution of the following vector-valued fractional delay differential equation

$$\dot{x}(t) = Ax(t) + Bx(t-\tau) - g\, I_2\, \mathrm{d}^{1-\alpha} x(t), \tag{32}$$

with

$$A = \frac{1}{2}\begin{bmatrix} -3 & 1 \\ 1 & -3 \end{bmatrix},\ B = \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},\ I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},\ \tau = 10,\ \alpha = \frac{5}{8}. \tag{33}$$

The matrices A and B are chosen so that (32) is asymptotically stable for any $g \geq 0$, $\tau \geq 0$, and $\alpha \in (0,1)$ [36, Thm. 7].

To study the stability of (32), we recast it into an abstract Cauchy problem

$$\frac{\mathrm{d}X}{\mathrm{d}t}(t) = \mathcal{A}X(t),\ X(0) \in H,\ X := \begin{pmatrix} x \\ \psi \\ \varphi \end{pmatrix} \in H, \tag{34}$$

which is known as an eigenvalue approach. The definition of $\mathcal{A}$ is obtained by using the diffusive representation of $\mathrm{d}^{1-\alpha}$ and rewriting the time-delay term as an observer of a transport equation on the bounded interval $(-\tau, 0)$ [14, § VI.6] [9, § 2.4] [34, Chap. 2], which leads to

$$\mathcal{A}X := \begin{pmatrix} Ax + B\psi(-\tau) - g\, I_2 \int_0^\infty [-\xi\varphi(\xi) + x]\, \mu_\alpha(\xi)\, \mathrm{d}\xi \\ \frac{\mathrm{d}\psi}{\mathrm{d}\theta} \\ -\xi\varphi(\xi) + x \end{pmatrix},$$

with state-space H and domain $\mathcal{D}(\mathcal{A})$ given by

$$H := \mathbb{C}^2 \times L^2(-\tau, 0; \mathbb{C}^2) \times L^2_{\xi\mu_\alpha(\xi)}(0, \infty; \mathbb{C}^2),$$

$$\mathcal{D}(\mathcal{A}) := \left\{ (x, \psi, \varphi) \in H \;\middle|\; \begin{array}{l} \psi \in H^1(-\tau, 0; \mathbb{C}^2) \\ -\xi\varphi(\xi) + x \in L^2_{(1+\xi)\mu_\alpha(\xi)}(0, \infty; \mathbb{C}^2) \end{array} \right\},$$

where the weighted L^2 spaces $L^2_{\xi\mu_\alpha(\xi)}$ and $L^2_{(1+\xi)\mu_\alpha(\xi)}$ are defined as

$$L^2_f(0,\infty) := \left\{ \varphi :\ (0,\infty) \to \mathbb{C} \text{ measurable} \;\middle|\; \int_0^\infty |\varphi(\xi)|^2 f(\xi)\, \mathrm{d}\xi < \infty \right\},$$

with $f(\xi) = \xi\mu_\alpha(\xi)$ and $f(\xi) = (1+\xi)\mu_\alpha(\xi)$, respectively. For additional background on this semigroup formulation, see [33,37].

Remark 11 *(Motivation).* The equation (32) is a toy model meant to check the suitability of a given discretization method for stability studies, thus validating its use for equations that do not enjoy theoretical results. Practical stability studies consist in computing stability regions, see e.g. [34] for delay equations. When using an eigenvalue approach, it is of paramount importance that the spectrum of $\mathcal{A}$ be accurately approximated, something which is less of concern with time-domain simulations.

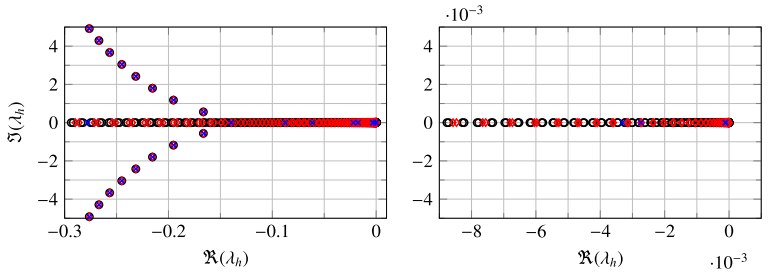

Fig. 13. Spectrum $\sigma(\mathcal{A}_h)$ for (32, 33) obtained with the $Q_{\beta,N}$ method with $\beta = \beta_2$. Transport equation discretized with $N_p = 80$ nodes. (○) $N = 400$ ($S(\mathcal{A}_h) = -8.12 \times 10^{-11}$). (◊) $N = 200$ ($S(\mathcal{A}_h) = -1.29 \times 10^{-9}$). (×) $N = 11$ ($S(\mathcal{A}_h) = -1.00 \times 10^{-4}$).

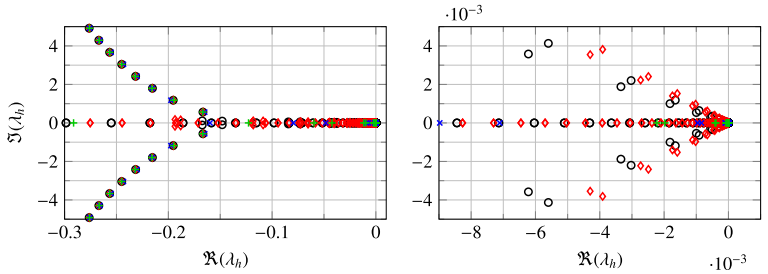

Fig. 14. Spectrum $\sigma(\mathcal{A}_h)$ for (32, 33) obtained with the optimization method with $\xi_{\max} = 10^4$. Transport equation discretized with $N_p = 80$ nodes. $N = 400$ with (○) $\xi_{\min} = 10^{-15}$ ($S(\mathcal{A}_h) = +3.5 \times 10^{12}$); (◊) $\xi_{\min} = 10^{-10}$ ($S(\mathcal{A}_h) = -7.1 \times 10^{-11}$). $N = 20$ with (×) $\xi_{\min} = 10^{-15}$ ($S(\mathcal{A}_h) = +2 \times 10^{-15}$); (+) $\xi_{\min} = 10^{-10}$ ($S(\mathcal{A}_h) = -10^{-12}$).

The stability of the fractional delay differential equation (32) follows from properties of the spectrum of $\mathcal{A}$. Theoretically, the spectrum of $\mathcal{A}$ consists of two distinct parts: (a) isolated eigenvalues with finite algebraic multiplicity; (b) an essential spectrum on $(-\infty, 0)$ if $g \neq 0$. This essential spectrum implies that

$$S(\mathcal{A}) := \sup_{\lambda \in \sigma(\mathcal{A})} \Re(\lambda) = 0,$$

so that (32) cannot be exponentially stable, but is indeed asymptotically stable.

Let us chose $g = 2$ and try to recover this stability result numerically, by computing the spectrum of $\mathcal{A}_h$, a finite-dimensional approximation of $\mathcal{A}$, which requires to discretize both the time-delay and the fractional derivative.

The monodimensional transport equation on $(-\tau, 0)$ can be discretized using any numerical scheme suited to the transport equation. Herein, we use a discontinuous Galerkin finite element method [22], whose spectral properties are well-known [23], on 1 element with N_p nodes (i.e. a polynomial of degree $N_p - 1$). For the large value $N_p = 80$, the spectrum is satisfactory in the region of interest, so that any witnessed spectral pollution stems from the approximation of the fractional derivative.

The fractional derivative is approximated with N variables φ_n, so that the matrix $\mathcal{A}_h$ is square with $(2 + N_p + N)$ lines. Figs. 13 and 14 plot the spectra obtained using both the $Q_{\beta,N}$ and the optimization methods. In both cases, the structure of the spectrum is consistent with the theory; there are however major differences between the two methods.

Since the $Q_{\beta,N}$ method has one parameter, it is straightforward to assess convergence. In the region of interest, the spectrum is converged for $N \geq 11$, see Fig. 13. The right plot of Fig. 13 shows that the essential spectrum is only made of real eigenvalues. Moreover, we have the most important property that

$$S(\mathcal{A}_h) := \max_{\lambda_h \in \sigma_p(\mathcal{A}_h)} \Re(\lambda_h)$$

is negative for all values of N. Hence, the stability result is verified numerically.

Let us now turn to the optimization method described in Section 4.2. Let us set $\xi_{\max} = 10^4$ so that the two remaining free parameters are N and the lower bound $\xi_{\min}$. Fig. 14 plots the spectra obtained using two values for N and $\xi_{\min}$, namely small ones ($N = 20$ and $\xi_{\min} = 10^{-15}$) and large ones ($N = 400$ and $\xi_{\min} = 10^{-10}$). The left plot shows that the structure of the spectrum is apparently identical to that obtained with the $Q_{\beta,N}$ method, with a reasonably converged point spectrum. However, the zoom given in the right plot shows that the essential spectrum is polluted. Significantly for a stability study, the spectrum can become slightly unstable, see the positive value of $S(\mathcal{A}_h)$ for $\xi_{\min} = 10^{-15}$: although $S(\mathcal{A}_h)$ remains close to zero, its sign depends on the choice of $\xi_{\min}$ and N. This implies that the optimization method is not suited to compute the spectrum of $\mathcal{A}$ in this example.

The Birk–Song method (27) also enjoys spectral accuracy, see Fig. 15. At $N = 11$, the spectrum is well-converged and the essential spectrum is both non-polluted and stable. This suggests the conjecture that spectral correctness is exhibited by

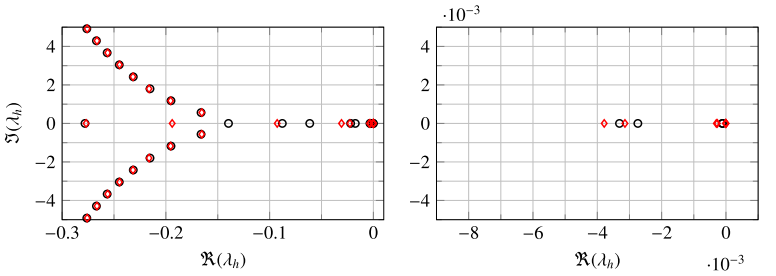

Fig. 15. Spectrum $\sigma(\mathcal{A}_h)$ for (32, 33) obtained with two quadrature methods with $N = 11$. Transport equation discretized with $N_p = 80$ nodes. (○) $Q_{\beta,N}$ method with $\beta = \beta_2$. ($S(\mathcal{A}_h) = -10^{-4}$). (◇) Birk–Song method (27) ($S(\mathcal{A}_h) = -3.5 \times 10^{-8}$).

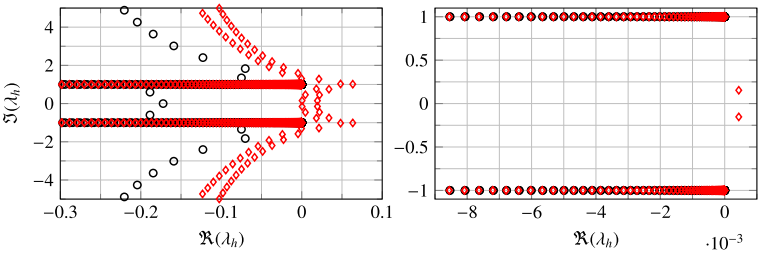

Fig. 16. Spectrum $\sigma(\mathcal{A}_h)$ for (35, 33) obtained with the $Q_{\beta,N}$ method with $\beta = \beta_2$ and $N = 400$. Transport equation discretized with $N_p = 80$ nodes. (○) $\tau_0 = 0$ ($S(\mathcal{A}_h) = -6 \times 10^{-10}$). (◇) $\tau_0 = 2\tau$ ($S(\mathcal{A}_h) = 6.3 \times 10^{-2}$).

every quadrature-based methods, so that they should be preferred to optimization-based ones for any application where a correct spectrum is needed.

Let us conclude this section with two additional examples.

Application to Bessel function. As recalled in Section 2.1, diffusive representations need not be restricted to fractional operators. Let us consider a more complex equation than (32), for instance the memory delay equation

$$\dot{x}(t) = Ax(t) + Bx(t-\tau) - g\, I_2\, J_0 \star x(t-\tau_0), \tag{35}$$

where $\tau_0 \geq 0$ and A, B, g, and τ are given by (33). Similarly to (32), the diffusive representation of J_0 given by (5) enables to formulate an abstract Cauchy problem (34). However, since the weight μ is complex-valued, the asymptotic stability of (35) cannot be established using the energy method followed in [36, Thm. 7] for (32). Hence the need for a numerical stability study. Fig. 16 plots the discrete spectrum obtained with the $Q_{\beta,N}$ method for two values of τ_0, namely $\tau_0 = 0$ and $\tau_0 = 2\tau$. The spectrum exhibits two straight lines that start from $-i$ and $+i$, which are the cuts chosen to extend the Laplace transform $\hat{J}_0$ to the left half-plane. This discretization enables us to conclude that the case $\tau_0 = 0$ is asymptotically stable while the case $\tau_0 = 2\tau$ is unstable.

Constrained optimization. The pollution of the essential spectrum visible in the right plot of Fig. 14 can be linked to the *unconstrained* nature of the linear least squares optimization described in Section 4.2. More specifically, it is caused by the negativity of some weights μ_n (which is not necessarily a practical concern for time-domain computations). An alternative is therefore to use, instead of the pseudo-inverse, an iterative optimization algorithm that enforces the nonnegativity constraint. This is illustrated by Fig. 17, which presents spectra obtained by minimizing (28) with $\mu \geq 0$ using the nonnegative least squares algorithm [24, (23.10)] through its implementation in MATLAB® `lsqnonneg`. The spectral pollution is significantly reduced, but has not completely disappeared, since slightly unstable spectra (or even unconverged spectra) can still be obtained for some poorly chosen parameters such as $\xi_{\min} = 10^{-15}$, $\xi_{\max} = 10^4$, and $N = 50$. The cost of this optimization algorithm, as well as its difficulty to converge for some triplets $(\xi_{\min}, \xi_{\max}, N)$, is a practical challenge to the computation of stability regions.

5. Improvement of the quadrature method using a nonlinear least squares optimization

The discussion of Section 4.3 has highlighted that the *sole* use of a nonlinear least squares minimization of the cost function (28) is not practical, due to both its computational cost and sensitivity to the initial pole distribution (i.e. the initial distribution of $(\xi_n)_n$). This section investigates the use of a nonlinear least squares minimization to refine the poles and weights given by the $Q_{\beta,N}$ method (see Definition 4), i.e. a *combined* use of an optimization and a quadrature rule. It emphasizes the importance of the cost function definition by comparing (37), (38), and (39) using a trust-region method;

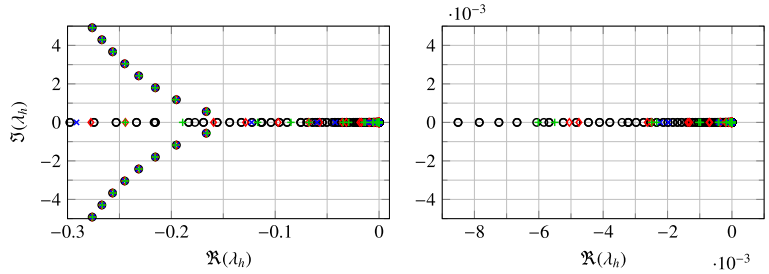

Fig. 17. Spectrum $\sigma(\mathcal{A}_h)$ for (32, 33) obtained by minimizing (28) with the nonnegativity constraint $\mu \geq 0$. Upper bound chosen as $\xi_{\max} = 10^4$. Transport equation discretized with $N_p = 80$ nodes. Lower bound $\xi_{\min} = 10^{-10}$ with (○) $N = 400$ ($S(\mathcal{A}_h) = -10^{-12}$); (◇) $N = 50$ ($S(\mathcal{A}_h) = -10^{-12}$); (×) $N = 20$ ($S(\mathcal{A}_h) = -10^{-12}$). (+) $N = 50$ with $\xi_{\min} = 10^{-15}$ ($S(\mathcal{A}_h) = +5.4 \times 10^{-16}$).

the numerical results show that the cost function (37) is to be preferred to build parsimonious approximations as it can deliver substantial improvements when the number of quadrature nodes is low.

Section 5.1 defines the numerical methodology as well as the covered cost functions, while Section 5.2 gathers the numerical results and concludes with practical guidelines. The MATLAB code is available online.[1]

5.1. Numerical methodology and considered cost functions

The purpose of the numerical methodology described below is to improve the poles and weights given by the $Q_{\beta,N}$ method by minimizing a given cost function J; from now on, this methodology is denoted $Q_{\beta,N}^{\text{OPT-}J}$. A similar methodology is used in [26, § 4.2] to improve the Birk–Song method (27).

Definition 12. The $Q_{\beta,N}^{\text{OPT-}J}$ discretization of (1) is (8) where the poles and weights are computed with the following three-step method.

1. Choose N (number of quadrature nodes), β (scalar parameter that sets the change of variable), and $(\xi, \mu) \mapsto J(\xi, \mu)$ (cost function).
2. Compute the poles $(\xi_n)_n$ and weights $(\mu_n)_n$ using the $Q_{\beta,N}$ method (see Definition 4).
3. Refine the computed poles and weights by minimizing J under the linear constraints

$$0 \leq \xi_n \quad \text{(a)} \qquad \xi_n \leq \xi_{\max} \quad \text{(b)} \qquad \mu_n \geq 0 \quad \text{(c)} \qquad (n \in \mathbb{N}), \tag{36}$$

starting with the values obtained in step 2.

The constraints (36) are motivated by the discussions of the previous sections: let us summarize their purposes. Condition (a) is required for stability, as it prevents any pole of $\hat{h}_{\text{num}}$ from having a nonnegative real part. Condition (b) ensures that the poles stay below the upper bound given by the $Q_{\beta,N}$ method, so that there is no time-step reduction with an explicit time-integration scheme. Condition (c) is optional but can be enforced when the diffusive weight $\xi \mapsto \mu(\xi)$ is nonnegative as it enables to get an unpolluted and stable spectrum (see Section 4.5 for an illustration of the impact of this constraint).

The use of a nonlinear least squares optimization implies an additional freedom in the definition of the cost function, compared with the linear least squares considered in Section 4.2. The three studied cost functions are formulated in the frequency domain. The first one is that already used in Section 4.2:

$$J(\xi, \mu) := \sum_{k=1}^{K} \left| \sum_{n=1}^{N} \frac{\mu_n}{i\omega_k + \xi_n} - \hat{h}(i\omega_k) \right|^2, \tag{37}$$

with the K angular frequencies logarithmically spaced in $[\xi_{\min}, \xi_{\max}]$

$$\omega_k = \xi_{\min} \left(\frac{\xi_{\max}}{\xi_{\min}} \right)^{\frac{k-1}{K-1}} \quad (k \in [\![1, K]\!]).$$

The second cost function is formulated so as to cancel an integrable singularity of $\omega \mapsto \hat{h}(i\omega)$ at $\omega = 0$ (consider $\hat{h} = \hat{Y}_\alpha$ with $\alpha \in (0, 1)$):

[1] https://github.com/fmonteghetti/Diffusive-Representation.

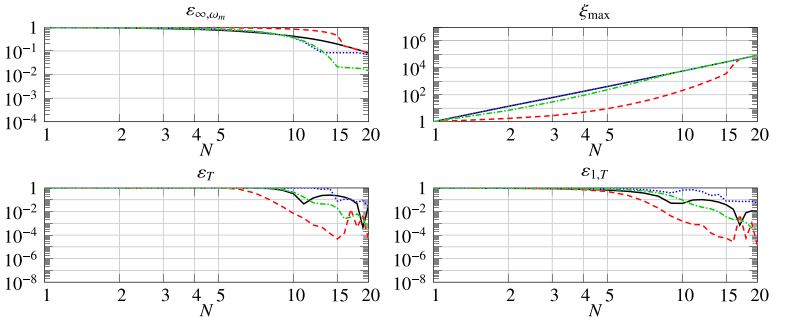

Fig. 18. Errors (23, 24, 25) and maximum pole (26) for $h = Y_\alpha$ with $\alpha = \frac{1}{2}$. (——) $Q_{\beta,N}$ with $\beta = \beta_2$. Quadrature rule combined with nonlinear least squares minimization using three different cost functions: (- - - -) $Q_{\beta_2,N}^{\mathrm{OPT}-J}$ (cost function given by (37), with tolerance relaxed to 10^{-14} for $N = 10$ and 17), (·····) $Q_{\beta_2,N}^{\mathrm{OPT}-J_s}$ (cost function (38)), (-·-·-) $Q_{\beta_2,N}^{\mathrm{OPT}-J_\psi}$ (cost function (39)).

$$J_s(\xi,\mu) := \sum_{k=1}^{K} \left| i\omega \sum_{n=1}^{N} \frac{\mu_n}{i\omega_k + \xi_n} - i\omega \hat{h}(i\omega_k) \right|^2. \tag{38}$$

The last cost function is more intricate:

$$J_\psi(\xi,\mu) := \sum_{k=1}^{K} \left| \psi\left(\sum_{n=1}^{N} \frac{\mu_n}{i\omega_k + \xi_n} \right) - \psi\left(\hat{h}(i\omega_k) \right) \right|^2, \tag{39}$$

where ψ is given by

$$\psi(s) := \frac{s-1}{s+1}. \tag{40}$$

Let us explain the reasoning behind the definition of the cost function J_ψ. Assume that $u \mapsto h \star u$ models a passive system [47]: the Laplace transform $\hat{h}$ is then a positive-real function [3, Thm. 3.15], the definition of which is recalled below, where

$$\mathbb{C}_0^+ = \{ s \in \mathbb{C} \,|\, \Re(s) > 0 \}.$$

Definition 13. A function $f : \mathbb{C}_0^+ \to \mathbb{C}$ is *positive-real* if (i) f is analytic in $\mathbb{C}_0^+$, (ii) $f(s) \in \mathbb{R}$ for $s \in (0,\infty)$, (iii) $\Re[f(s)] \geq 0$ for $s \in \mathbb{C}_0^+$.

The definition of ψ in (39) is then suggested by the fact that ψ is a conformal map of the open half-plane $\mathbb{C}_0^+$ onto the open unit disk [15, Chap. XI]. This transformation is linked to the so-called scattering formulation [3,27].

5.2. Numerical results

This section gathers numerical comparisons between the $Q_{\beta,N}$ and $Q_{\beta,N}^{\mathrm{OPT}\text{-}J}$ methods for the three cost functions introduced above, namely (37), (38), and (39).

The minimization is done using a subspace trust-region method based on the interior-reflective Newton method described in [7,8] through its implementation in MATLAB® `lsqnonlin`. The analytical expression of the jacobian matrix $\nabla_{\xi,\mu} J(\xi,\mu)$ is supplied to the algorithm. Unless explicitly mentioned, all the termination tolerances are set to 10^{-15}. For the numerical comparisons we use the same parameters as in the previous sections, namely (25, 29). (Reducing K does make convergence easier to attain but worsens the results.)

Method $Q_{\beta,N}^{\mathrm{OPT}-J}$ based on the cost function (37). Fig. 18 shows the comparison with the $Q_{\beta,N}$ method for the approximation of the fractional kernel (3) with $\alpha = 1/2$. The top left graph shows that the $Q_{\beta,N}^{\mathrm{OPT}-J}$ method does not reduce the error in $L^\infty(-\omega_n,\omega_n)$, which may be justified by the fact that it relies on a least squares optimization in $[\xi_{\min},\xi_{\max}]$. Let us now comment the three other graphs, which are more relevant for time-domain simulations. For a moderate number of variables, this method yields a significant improvement which can be seen on the reduction of both the errors and maximum pole $\xi_{\max}$. This improvement can be noticed when solving the FDE (30), as illustrated in the right graph of Fig. 19, which plots the relative error when solving with $N = 6$: not only is the error reduced compared to the $Q_{\beta,N}$ method but the maximum stable time step is also increased, since $\xi_{\max}$ is lower. Here, the $Q_{\beta,N}^{\mathrm{OPT}-J}$ method is superior to the $Q_{\beta,N}$ method. However, as N increases past $N = 10$ (20 optimization variables), the convergence of the trust-region method becomes more difficult to achieve and the improvements fade away. This can be seen on the right graph of Fig. 20, where for $N = 15$ the accuracy delivered by the $Q_{\beta,N}^{\mathrm{OPT}-J}$ method is worse than that of the $Q_{\beta,N}$ method.

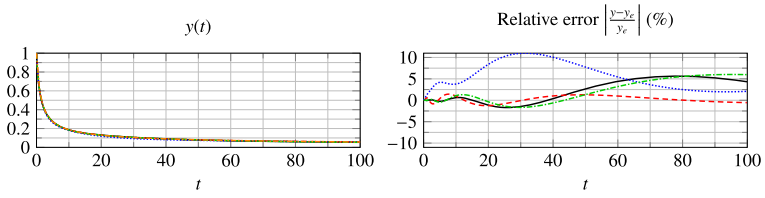

Fig. 19. FDE (30) for $y_0 = 1$ and $g = 1$. Numerical solutions computed with RKF84, $\Delta t = 9 \times 10^{-3}$, and $N = 6$. (——) $Q_{\beta,N}$ with $\beta = \beta_2$ ($\xi_{\min} = 1.221 \times 10^{-3}$, $\xi_{\max} = 8.189 \times 10^{2}$). (- - - -) $Q_{\beta_2,B}^{\mathrm{OPT}-J}$ ($\xi_{\min} = 2.960 \times 10^{-4}$, $\xi_{\max} = 1.747 \times 10^{1}$). (·····) $Q_{\beta_2,B}^{\mathrm{OPT}-J_s}$ ($\xi_{\min} = 5.920 \times 10^{-3}$, $\xi_{\max} = 8.189 \times 10^{2}$). (-·-·-) $Q_{\beta_2,B}^{\mathrm{OPT}-J_\psi}$ ($\xi_{\min} = 2.018 \times 10^{-3}$, $\xi_{\max} = 5.371 \times 10^{2}$). (Left only) (-··-··) Exact solution (31).

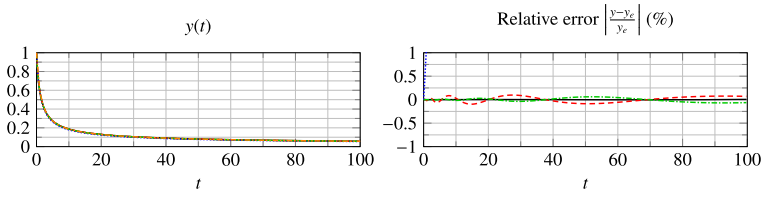

Fig. 20. FDE (30) for $y_0 = 1$ and $g = 1$. Numerical solutions computed with RKF84, $\Delta t = 2.8 \times 10^{-4}$ and $N = 15$. (——) $Q_{\beta,N}$ with $\beta = \beta_2$ ($\xi_{\min} = 3.648 \times 10^{-5}$, $\xi_{\max} = 2.741 \times 10^{4}$). (- - - -) $Q_{\beta_2,B}^{\mathrm{OPT}-J}$ ($\xi_{\min} = 3.417 \times 10^{-6}$, $\xi_{\max} = 3.776 \times 10^{3}$). (·····) $Q_{\beta_2,B}^{\mathrm{OPT}-J_s}$ ($\xi_{\min} = 3.696 \times 10^{-5}$, $\xi_{\max} = 2.741 \times 10^{4}$, maximum error 11%). (-·-·-) $Q_{\beta_2,B}^{\mathrm{OPT}-J_\psi}$ ($\xi_{\min} = 1.815 \times 10^{-5}$, $\xi_{\max} = 2.741 \times 10^{4}$). (Left only) (-··-··) Exact solution (31).

Method $Q_{\beta,N}^{\mathrm{OPT}-J_s}$ based on the cost function (38). The trust-region method converges slightly faster with J_s than with J and it delivers a better error in the frequency domain, as can be seen on the top left graph of Fig. 18. However, the two bottom graphs show that the time-domain errors are worsened. The time-domain simulation results presented in Figs. 19 and 20 indicates that all the other methods outperform the $Q_{\beta,N}^{\mathrm{OPT}-J_s}$ method: J_s is therefore not an improvement over J, at least with the employed trust-region algorithm.

Method $Q_{\beta,N}^{\mathrm{OPT}-J_\psi}$ based on the cost function (39). The cost function J_ψ yields the fastest convergence of the trust-region algorithm, especially for a large number of quadrature nodes: this justifies a posteriori why this cost function has been introduced above. The top left graph of Fig. 18 shows that a significant improvement is obtained in the frequency domain. By contrast, the time-domain errors can be larger than those obtained with J for small values of N, but improve as N increases. Let us now look at the errors obtained when solving the FDE (30) with $N = 6$ and $N = 10$, plotted in Figs. 19 and 20, respectively. For $N = 6$, the $Q_{\beta,N}^{\mathrm{OPT}-J_\psi}$ method is roughly equivalent to the $Q_{\beta,N}$ method, with no significant improvement or worsening of the error. For $N \geq 10$, the reduction in the approximation errors seen on Fig. 18 does not necessarily translate as a reduced error in the solution of the FDE. For example, for $N = 15$, the $Q_{\beta,N}^{\mathrm{OPT}-J_\psi}$ method delivers a worse solution than the $Q_{\beta,N}$ method, although it fares better than the $Q_{\beta,N}^{\mathrm{OPT}-J}$ method.

Summary. The following guidelines can be deduced from the numerical results. The $Q_{\beta,N}$ method is a convenient choice for time-domain simulations, as it delivers accurate results without the need for a nonlinear least squares optimization whose convergence is not guaranteed. If the application at hand mandates a low number of quadrature nodes, then the $Q_{\beta,N}^{\mathrm{OPT}-J}$ method should yield satisfactory results. The main interest of the $Q_{\beta,N}^{\mathrm{OPT}-J_\psi}$ method is that it scales better than $Q_{\beta,N}^{\mathrm{OPT}-J}$ with respect to N, with swift convergence and no dramatic worsening of the initial poles and weights; however, it may not deliver a significant improvement.

6. Conclusion and outlook

This paper has focused on the discretization using Gaussian quadrature of a class of diffusive kernels that contains the zeroth-order Bessel function of the first kind (5) as well as the fractional kernel (3).

Section 2 has introduced the proposed $Q_{\beta,N}$ discretization method and Section 3 has shown that β must be tailored to the kernel at hand. The choice $\beta = \min(\alpha, 1-\alpha)$, where α is the power of the diffusive weight singularity, has proven to be a suitable choice for both time-domain computations and eigenvalue problems. In particular, for rational α, a spectral convergence rate can be attained. Numerical comparison with an existing discretization method based on the Gauss–Jacobi quadrature rule has shown the complementarity of the two approaches: although the $Q_{\beta,N}$ method can have a slower convergence, see Figs. 6 and 7, it is more suited to numerical simulations using an explicit scheme due to its lower $\xi_{\max}$, see Section 4.4.

However, both methods are spectrally correct, in the sense that they yield an unpolluted and convergent approximation of the essential spectrum (linked to the diffusive operator), by contrast with optimization-based methods that can yield

polluted spectra whose convergence is difficult to assess. This property, highlighted on the fractional delay differential equation (32), should be verified with any quadrature-based discretization method, so that quadrature-based methods should be preferred to optimization-based ones for any application where a correct spectrum is needed, for instance eigenvalue problems that arise in fractional ordinary and partial differential equations.

Nonetheless, this does not mean that optimization-based methods should be ignored. Section 5 has shown that for time-domain simulations, combining a nonlinear least squares minimization of (37) with the $Q_{\beta,N}$ method can yield substantial improvements for a low number of quadrature nodes N, convergence becoming an issue for a large number of nodes. Although the alternative cost function (39) has been shown to scale better with N, it has not delivered significant improvements with the employed trust-region algorithm.

Another important feature of optimization-based methods is their flexibility, which makes them applicable to a large class of diffusive operators encountered e.g. in wave propagation problems. Indeed, the main limitation of the proposed $Q_{\beta,N}$ method, as well as other existing quadrature-based discretization method, is that it does not apply to diffusive kernels whose diffusive weight μ is less well-behaved, for instance with sharp variations or oscillations. An example is

$$\hat{h}(s) = \frac{e^{-\epsilon\sqrt{s}}}{1-\rho e^{-2(s+\epsilon\sqrt{s})}}$$

with $\rho \in (-1,1)$ and $\epsilon > 0$ with $\epsilon \ll 1$, which is encountered with the Webster–Lokshin equation [30, Chap. 6] and impedance models of sound absorbing materials [38, § 6] [35, Chap. 2]. Another example encountered in acoustics is [38, § 5]

$$\hat{h}(s) = \frac{1}{a_0 + a_{\frac{1}{2}}\sqrt{s} + a_1 s},$$

where a_0, $a_{\frac{1}{2}}$, and a_1 are positive coefficients such that $a_1 \gg a_{\frac{1}{2}}$.

In both cases, any method that solely relies on change of variables and quadrature rules breaks down due to the fact that the corresponding diffusive weight μ can have a singular or near-singular behavior *within* its domain $(0,\infty)$. Circumventing this issue would first require locating these (near)-singularities. Existing adaptive quadrature algorithms, possibly combined with splitting the integration domain, may provide a satisfactory answer; for instance, in most cases, MATLAB® `integral` [43] is able to accurately compute $\int e^{-\xi t}\mu(\xi)\mathrm{d}\xi$ for $t > 0$.

Acknowledgement

This research has been financially supported by the French ministry of defense (Direction Générale de l'Armement) and ONERA (the French Aerospace Lab).

References

[1] K. Atkinson, An Introduction to Numerical Analysis, 2nd ed., John Wiley & Sons, New York, 1989.
[2] J. Baranowski, Quadrature based approximations of non-integer order integrator on finite integration interval, in: A. Babiarz, A. Czornik, J. Klamka, M. Niezabitowski (Eds.), Theory and Applications of Non-Integer Order Systems, Springer International Publishing, Cham, 2017, pp. 11–20.
[3] E.J. Beltrami, M.R. Wohlers, Distributions and the Boundary Values of Analytic Functions, Academic Press, New York, 1966.
[4] C. Birk, C. Song, An improved non-classical method for the solution of fractional differential equations, Comput. Mech. 46 (5) (2010) 721–734, https://doi.org/10.1007/s00466-010-0510-4.
[5] M. Caputo, Vibrations of an infinite plate with a frequency independent Q, J. Acoust. Soc. Am. 60 (3) (1976) 634–639, https://doi.org/10.1121/1.381126.
[6] C. Casenave, G. Montseny, Introduction to diffusive representation, in: 4th IFAC Symposium on System, Structure and Control, Ancona, Italy, 2010.
[7] T.F. Coleman, Y. Li, On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds, Math. Program. 67 (1) (1994) 189–224, https://doi.org/10.1007/BF01582221.
[8] T. Coleman, Y. Li, An interior trust region approach for nonlinear minimization subject to bounds, SIAM J. Optim. 6 (2) (1996) 418–445, https://doi.org/10.1137/0806023.
[9] R.F. Curtain, H. Zwart, An Introduction to Infinite-Dimensional Linear Systems Theory, Springer, New York, 1995.
[10] P. Davis, P. Rabinowitz, Methods of Numerical Integration, 2nd ed., Academic Press, San Diego, 1984.
[11] W. Desch, R.K. Miller, Exponential stabilization of Volterra integral equations with singular kernels, J. Integral Equ. Appl. 1 (3) (1988) 397–433, https://doi.org/10.1216/JIE-1988-1-3-397.
[12] K. Diethelm, An investigation of some nonclassical methods for the numerical approximation of Caputo-type fractional derivatives, Numer. Algorithms 47 (4) (2008) 361–390, https://doi.org/10.1007/s11075-008-9193-8.
[13] J. Dunham, The Theory of Approximation, American Mathematical Society, New York, 1930.
[14] K.-J. Engel, R. Nagel, One-Parameter Semigroups for Linear Evolution Equations, Springer-Verlag, New York, ISBN 0-387-98463-1, 2000.
[15] T.W. Gamelin, Complex Analysis, Springer-Verlag, New York, ISBN 0-387-95093-1, 2001.
[16] G. Garcia, J. Bernussou, Identification of the dynamics of a lead acid battery by a diffusive model, in: ESAIM: Proceedings, vol. 5, EDP Sciences, 1998, pp. 87–98.
[17] R. Garrappa, Numerical evaluation of two and three parameter Mittag-Leffler functions, SIAM J. Numer. Anal. 53 (3) (2015) 1350–1369, https://doi.org/10.1137/140971191.
[18] R. Garrappa, F. Mainardi, M. Guido, Models of dielectric relaxation based on completely monotone functions, Fract. Calc. Appl. Anal. 19 (5) (2016) 1105–1160, https://doi.org/10.1515/fca-2016-0060.
[19] G. Gripenberg, S.-O. Londen, O.J. Staffans, Volterra Integral and Functional Equations, Cambridge University Press, Cambridge, ISBN 0-521-37289-5, 1990.

[20] H. Haddar, J.-R. Li, D. Matignon, Efficient solution of a wave equation with fractional-order dissipative terms, Am. J. Comput. Appl. Math. 234 (6) (2010) 2003–2010, https://doi.org/10.1016/j.cam.2009.08.051.
[21] T. Hélie, D. Matignon, Representations with poles and cuts for the time-domain simulation of fractional systems and irrational transfer functions, Signal Process. 86 (10) (2006) 2516–2528, https://doi.org/10.1016/j.sigpro.2006.02.017.
[22] J.S. Hesthaven, T. Warburton, Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications, Springer, New York, 2008.
[23] F.Q. Hu, M. Hussaini, P. Rasetarinera, An analysis of the discontinuous Galerkin method for wave propagation problems, J. Comput. Phys. 151 (2) (1999) 921–946, https://doi.org/10.1006/jcph.1999.6227.
[24] C. Lawson, R. Hanson, Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, New Jersey, 1974.
[25] J.-R. Li, A fast time stepping method for evaluating fractional integrals, SIAM J. Sci. Comput. 31 (6) (2010) 4696–4714, https://doi.org/10.1137/080736533.
[26] B. Lombard, D. Matignon, Diffusive approximation of a time-fractional Burger's equation in nonlinear acoustics, SIAM J. Appl. Math. 76 (5) (2016) 1765–1791, https://doi.org/10.1137/16M1062491.
[27] R. Lozano, B. Brogliato, O. Egeland, B. Maschke, Dissipative Systems Analysis and Control: Theory and Applications, Springer-Verlag, London, 2000.
[28] C. Lubich, Discretized fractional calculus, SIAM J. Math. Anal. 17 (3) (1986) 704–719, https://doi.org/10.1137/0517050.
[29] F. Mainardi, Fractional Calculus: Some Basic Problems in Continuum and Statistical Mechanics, Springer Verlag, 1997.
[30] D. Matignon, Représentations en variables d'état de modèles de guides d'ondes avec dérivation fractionnaire, Ph.D. thesis, Université Paris XI Orsay, 1994.
[31] D. Matignon, Stability properties for generalized fractional differential systems, in: ESAIM: Proceedings, vol. 5, EDP Sciences, 1998, pp. 145–158.
[32] D. Matignon, An introduction to fractional calculus, in: P. Abry, P. Gonçalvès, J. Levy-Vehel (Eds.), Scaling, Fractals and Wavelets, ISTE–Wiley, London–Hoboken, ISBN 978-1-84-821072-1, 2009, pp. 237–277.
[33] D. Matignon, C. Prieur, Asymptotic stability of Webster–Lokshin equation, Math. Control Relat. Fields 4 (4) (2014) 481–500, https://doi.org/10.3934/mcrf.2014.4.481.
[34] W. Michiels, S.-I. Niculescu, Stability, Control, and Computation for Time-Delay Systems, 2nd ed., SIAM, Philadelphia, ISBN 978-0-898716-32-0, 2014.
[35] F. Monteghetti, Analysis and Discretization of Time-Domain Impedance Boundary Conditions in Aeroacoustics, Ph.D. thesis, ISAE-SUPAERO, Université de Toulouse, Toulouse, France, 2018.
[36] F. Monteghetti, G. Haine, D. Matignon, Stability of linear fractional differential equations with delays: a coupled parabolic-hyperbolic PDEs formulation, in: 20th World Congress of the International Federation of Automatic Control, IFAC, 2017.
[37] F. Monteghetti, G. Haine, D. Matignon, Asymptotic stability of the multidimensional wave equation coupled with classes of positive real impedance boundary conditions, submitted for publication, 2018.
[38] F. Monteghetti, D. Matignon, E. Piot, L. Pascal, Design of broadband time-domain impedance boundary conditions using the oscillatory-diffusive representation of acoustical models, J. Acoust. Soc. Am. 140 (3) (2016) 1663–1674, https://doi.org/10.1121/1.4962277.
[39] G. Montseny, Diffusive representation of pseudo-differential time-operators, in: ESAIM: Proceedings, vol. 5, EDP Sciences, 1998, pp. 159–175.
[40] I. Podlubny, Fractional Differential Equations, Academic Press, San Diego, 1999.
[41] S.G. Samko, A.A. Kilbas, O.I. Marichev, Fractional Integrals and Derivatives, Gordon and Breach, Yverdon, Switzerland, 1993.
[42] R. Scherer, S.L. Kalla, Y. Tang, J. Huang, The Grünwald–Letnikov method for fractional differential equations, Comput. Math. Appl. 62 (3) (2011) 902–917, https://doi.org/10.1016/j.camwa.2011.03.054.
[43] L. Shampine, Vectorized adaptive quadrature in MATLAB, Am. J. Comput. Appl. Math. 211 (2) (2008) 131–140, https://doi.org/10.1016/j.cam.2006.11.021.
[44] O.J. Staffans, Well-posedness and stabilizability of a viscoelastic equation in energy space, Trans. Am. Math. Soc. 345 (2) (1994) 527–575, https://doi.org/10.1090/S0002-9947-1994-1264153-X.
[45] T. Toulorge, W. Desmet, Optimal Runge–Kutta schemes for discontinuous Galerkin space discretizations applied to wave propagation problems, J. Comput. Phys. 231 (4) (2012) 2067–2091, https://doi.org/10.1016/j.jcp.2011.11.024.
[46] L. Yuan, O. Agrawal, A numerical scheme for dynamic systems containing fractional derivatives, ASME J. Vib. Acoust. 124 (2) (2002) 321–324, https://doi.org/10.1115/1.1448322.
[47] A. Zemanian, Distribution Theory and Transform Analysis, McGraw-Hill, 1965.